Published in final edited form as:

Nature. 2021 May ; 593(7858): 294–298. doi:10.1038/s41586-021-03458-7.

Structural basis of Long-range to Short-range synaptic transition in NHEJ

Siyu Chen^{1,2}, Linda Lee³, Tasmin Naila⁴, Susan Fishbain¹, Annie Wang¹, Alan E Tomkinson⁴, Susan P Lees-Miller³, Yuan He^{1,2,5,6,*}

¹Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA

²Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, USA

³Department of Biochemistry and Molecular Biology, Robson DNA Science Centre and Charbonneau Cancer Institute, University of Calgary, Alberta, T2N 4N1, Canada

⁴Department of Internal Medicine and Molecular Genetics & Microbiology and the University of New Mexico Comprehensive Cancer Center, University of New Mexico, Albuquerque, New Mexico 87131

⁵Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, USA

⁶Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Northwestern University, Chicago, USA

Abstract

DNA double-strand breaks (DSBs) are the most cytotoxic form of DNA damage, with their aberrant repair linked with carcinogenesis^{1,2}. The conserved error-prone Non-Homologous End-Joining (NHEJ) pathway plays a key role in determining the effects of DSB-inducing agents used to treat cancer as well as the generation of antibody and T cell receptor diversity^{2,3}. Here, we applied single-particle cryo-electron microscopy (EM) to visualize two key DNA-protein complexes formed by NHEJ factors. Ku, DNA-PKcs, LigIV-XRCC4, and XLF form a Long-range synaptic complex, in which the DNA ends are held ~115 Å apart. Two DNA end-bound Ku-DNA-PKcs subcomplexes are linked by DNA-PKcs-DNA-PKcs interactions and a LigIV-XRCC4-XLF-XRCC4-LigIV scaffold. The relative orientation of the DNA-PKcs molecules suggests a mechanism for auto-phosphorylation *in trans*, leading to dissociation of DNA-PKcs and transition into the Short-range synaptic complex. Within this complex, the Ku-bound DNA ends are aligned for processing and ligation by the XLF-anchored scaffold, and a single LigIV catalytic domain is stably associated with a nick between the two Ku molecules, suggesting that joining of both strands of a DSB involves both LigIV molecules.

*Corresponding author (yuanhe@northwestern.edu).

Author contributions

Y He, SP Lees-Miller, and AE Tomkinson conceived the project. S Chen performed most of the experiments and collected and analyzed cryo-EM data with Y He. L Lee and T Naila contributed to protein purification. S Chen built the models with help from Y He. S Chen, S Fishbain, A Wang, and Y He wrote the manuscript with input from all other authors.

Competing interests

There are no competing interests.

Cytotoxic and mutagenic DSBs are repaired rapidly by NHEJ in proliferating and non-dividing cells². To initiate NHEJ, the Ku70/80 heterodimer (Ku) binds to a DNA end, recruiting the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), a phosphoinositide 3-kinase-related (PIKK) family kinase to form the DNA-PK^{4,5}. Additional core factors, DNA ligase IV (LigIV)-XRCC4 and XLF, are independently recruited to Ku-bound ends^{6,7}. Interestingly, each of the core NHEJ factors has end-bridging activity⁸⁻¹². DNA end engagement activates DNA-PKcs activity¹³, with auto-phosphorylation *in trans* resulting in its release from DSB ends^{14,15}. There are, however, contrasting reports as to whether DNA-PKcs is required for DNA end-bridging during NHEJ¹⁶⁻¹⁹. Furthermore, while there are partial structures of the core NHEJ components^{5,20-25}, how these factors coordinate to bring DNA ends together is unclear. Here we determine the structures of two key NHEJ complexes; a Long-Range (LR) synaptic complex that brings two DNA ends into proximity and a Short-Range (SR) synaptic complex in which the DNA ends are aligned for ligation.

Architecture of the LR synaptic complex

While co-incubation of Ku, DNA-PKcs, LigIV-XRCC4, and XLF^{19,26} with duplex DNA substrates (Extended Data Fig. 1a, Methods) yielded only DNA-PK complexes (Extended Data Fig. 1a-c), pre-assembly of DNA-PK on DNA ends followed by incubation with LigIV-XRCC4 and XLF generated larger complexes (Extended Data Fig. 1d-f). Using this order of addition approach and DNA substrates optimized for complex assembly (Extended Data Fig. 1g-i), we determined the structure of this larger complex to 4.6 Å resolution using C2 symmetry (Fig. 1a, Extended Data Fig. 1j-l, Extended Data Fig. 2). To improve resolution, symmetry expansion was applied to combine the signal from both halves of the complex. Signal subtraction followed by focused 3D classification and refinement revealed a 4.1 Å resolution reconstruction of DNA-PK complexed with the LigIV N-terminal BRCT (N-BRCT) domain. A similar focused classification resulted in a 9.3 Å reconstruction of the flexible XLF-XRCC4-LigIV sub-complex. With these density maps, we built a full model of this ~1.66 MDa complex that contains two copies of DNA-PK and LigIV-XRCC4 (one LigIV complexed with an XRCC4 homodimer) on opposite sides. In accord with smFRET studies with *Xenopus* egg extracts^{27,28}, a single XLF homodimer bridges the two halves (Fig. 1b, Supplementary Video 1). While the complex contains two DNA molecules, the ends are ~115 Å apart (Fig. 1c), suggesting that this LR synaptic complex corresponds to a similar complex identified by smFRET²⁶.

Multiple interaction surfaces contribute to the stabilization of the LR complex (Extended Data Fig. 1m). As in the XLF Ku-Binding Motif (KBM)-bound Ku crystal structure²⁹, the Ku80 von Willebrand (vWA) domain is rotated outward compared to the unbound state (Extended Data Fig. 3a-d). Similarly, the Ku70 vWA domain adopts an open state while engaging DNA-PKcs^{5,30}. While the head-to-head interaction surface between XLF and XRCC4 dimers is similar to that observed in the filamentous structure³¹⁻³³, the angle between the XLF and XRCC4 coiled-coil (CC) domains is larger (~30°) in the LR complex compared with these structures (Extended Data Fig. 4a), suggesting a disfavored configuration to accommodate linking of the two DNA-PK complexes. While XLF-XRCC4 filaments³¹⁻³³ may reduce DNA flexibility adjacent to the DSB, the LR complex tethers the

DSB ends. The XRCC4 CC domain is also bent by $\sim 15^\circ$ more compared to the CC domain in LigIV-XRCC4 alone (Extended Data Fig. 4b), again likely accommodating linking of the two DNA-PK complexes by the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold. Finally, the conformation of LigIV-XRCC4 in the LR complex is more similar to the *S. cerevisiae* Lig4p-Lif1p crystal structure than human LigIV-XRCC4³⁴, indicating the LigIV BRCT-XRCC4 CC interface is flexible (Extended Data Fig. 4c).

Three novel inter-subunit interactions were revealed within the LR complex. First, there is an interaction between conserved surfaces on N-BRCT and the Ku core region (Fig. 1d, Extended Data Fig. 5a), predicted by biochemical pull-down experiments³⁵. Secondly, there is an extra loop within a conserved, negatively-charged groove formed by the DNA-PKcs FAT domain²⁰ (Extended Data Fig. 5c) that we attribute to the conserved, positively-charged region (Supplementary Figure 5) within XRCC4 C-terminal tails. These interactions likely stabilize contacts between the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold and the DNA-PK complexes and guide XRCC4 C-terminal phosphorylation sites³⁶ to the DNA-PKcs active site. Notably, XRCC4 truncation mutation, R275X (X: stop codon), associated with both prenatal and postnatal growth failure and leukopenia², and identified in the cancer mutation database³⁷, is likely to disrupt the XRCC4-DNA-PKcs interaction and XRCC4 phosphorylation. Thirdly, we identified two conserved loops in the DNA-PKcs M-HEAT domain at the DNA-PK dimer interface (Fig. 1e) with loop 2569-2585 interacting with the highly conserved YRPD motif³⁸ (Supplementary Figure 1 and Extended Data Fig. 5b). Both loops, together with the YRPD motif, likely serve as a molecular sensor to detect the close juxtaposition of the two DNA-PK complexes.

***In trans* DNA-PKcs auto-phosphorylation**

While similar overall, superimposition of DNA-PK within the LR complex with DNA-PK alone^{30,39} revealed significant differences (Fig. 2a). In the LR complex, the DNA-PKcs head module, including the FAT and the kinase domains, is rotated by $\sim 10^\circ$ to further expose the catalytic center, a conformation that aligns better with the recent structure determined by Chen *et al.*³⁹. The Ku core region, together with the associated DNA-PKcs N-HEAT domain in the LR complex, is also shifted by ~ 4 Å, threading the DNA end further into the cradle of DNA-PKcs by ~ 2 bp (Fig. 2a,b). Moreover, a 32-amino acid DNA end-blocking (DEB) helix (residues 2736-2767) that spans the large empty space cradled by the DNA-PKcs HEAT repeats is stabilized. This helix, together with the linker loop connecting it and the DEB-appended (DEB-A) helix (residues 2724-2730), makes direct contact with the DNA end, disrupting end base-pairing (Fig. 2b). Interestingly, the DEB helix is flanked by the unstructured ABCDE auto-phosphorylation cluster¹⁴ and the conserved YRPD motif, suggesting that it coordinates DNA-PKcs dimer interactions and autophosphorylation of the ABCDE cluster (Extended Data Fig. 5d).

We compared the active sites of DNA-PK alone and within the LR complex with those of mTOR⁴⁰ and CDK2⁴¹ (Fig. 2c). The PIKK regulatory domain (PRD), which negatively regulates kinase activity by occupying the putative substrate-binding groove^{20,40}, is structured in both structures of DNA-PK alone^{30,39}, but mostly disordered in intrinsically active mTOR kinase⁴⁰, and completely disordered in the LR complex. Furthermore, the

orientation between the two DNA-PKcs molecules differs substantially from the self-inhibitory conformations of the ATM and the ATR-ATRIP dimer structures^{42,43} (Extended Data Fig. 6). Thus, DNA-PKcs in the LR complex is likely to be active. Intriguingly, both the ABCDE and the PQR clusters of one DNA-PKcs molecule are closer to the catalytic center of the other one (Fig. 2d), favoring *in trans* autophosphorylation¹⁴.

DNA-PKcs autophosphorylation has been hypothesized to destabilize binding to DNA ends, triggering DNA-PKcs dissociation from repair loci¹⁵. In our model, the ABCDE and PQR clusters are in close proximity to other phosphorylation sites at the DNA-PKcs dimer interface, suggesting that autophosphorylation could function as a substantial and coordinated electrostatic switch (Fig. 2d). With increased negative charge due to phosphorylation, the ABCDE cluster could compete with DNA for the DNA binding groove near the DEB helix (Fig. 2b, Extended Data Fig. 5d), triggering the dissociation of DNA-PKcs from the DNA end and the transition from the LR to the SR synaptic state.

Architecture of the SR synaptic complex

Although the LR complex structure reveals how canonical NHEJ factors initially tether DNA ends, they are not oriented for joining, and the LigIV catalytic domains are not visible, presumably due to their flexibility⁴⁴. In the absence of DNA PKcs, complexes were assembled on DNA substrates with either complementary or blunt ends (Extended Data Fig. 7a-h, Methods). After substrate optimization for complex stability (Extended Data Fig. 7i-l, Methods) and by collecting a large dataset, we were able to improve the resolution to 8.4 Å for the overall reconstruction and 6.8-7.8 Å resolution for different bodies after focused refinements, clearly revealing secondary structures of both NHEJ proteins and the DNA substrates (Fig. 3a, Extended Data Fig. 7m-n, Extended Data Fig. 8). With this density map, we directly docked in the subunit models from the LR complex as rigid bodies, revealing a very similar overall architecture to the LR complex, except for the two DNA-PKcs molecules. In accord with smFRET studies using *Xenopus* egg extracts^{27,28}, Ku and LigIV-XRCC4 subcomplexes reside on opposite sides of the complex, with one XLF dimer bridging the two halves (Fig. 3b, Supplementary Video 2). A perfectly aligned duplex DNA is visible between two Ku rings located at opposite ends of the complex, suggesting the captured species corresponds to the SR complex described previously²⁶. In our SR complex, a single LigIV catalytic domain is visible, engaging a single non-ligatable nick in the middle of the complex (Extended Data Fig. 7i).

Interactions observed in the LR complex, including XLF-XRCC4, XLF-Ku, LigIV-XRCC4, as well as the new LigIV N-BRCT-Ku interface, also occur in the SR complex and are mostly unchanged (Extended Data Fig. 3e,f, 4c, 7o). Interestingly, the angle between XLF and XRCC4 CCs on either end of the complex decreases to ~20°, partially restoring the fully relaxed states observed in several crystal structures³¹⁻³³ (Extended Data Fig. 4a). Concurrently, the gradual bending of the long XRCC4 CC is restored to 160° for the two protomers, much closer to the ground state observed in the structures of LigIV-XRCC4 alone (Extended Data Fig. 4b)²⁴.

New interactions in the SR complex include an interaction surface between the LigIV DNA binding domain (DBD) and the Ku70 vWA domain within the Ku ring on the opposite side of the complex that partially overlaps with the Ku70 interface with DNA-PKcs (Fig. 3c, Extended Data Fig. 9a). This explains why grouping the LigIV catalytic domain with its interacting Ku improves the resolution of both parts (Extended Data Fig. 8). It also demonstrates that, when engaging a DNA nick, a single LigIV molecule contacts both Ku molecules via independent interactions with the LigIV N-BRCT and DBD, providing insights as to how Ku stimulates DNA joining by LigIV⁶. The SR complex is further stabilized by interactions between XLF and the two Ku70 molecules (Fig. 3d). This is consistent with the reduced ability of XLF with amino acid substitutions (L174A, R178A, and L179A) in these interfaces to stimulate NHEJ²³ and may provide insights into the D166R cancer mutation³⁷ associated with immunodeficiency, microcephaly, and radiosensitivity². Thus, amino acid changes within a conserved patch near the C-terminal tip of XLF CC likely exert their effects by disrupting XLF-Ku70 interactions within the SR complex (Extended Data Fig. 9b).

Tandem ligation by two LigIV molecules

Even with substrates containing two nicks, only a single LigIV catalytic domain was visible within the SR complex (Extended Data Fig. 7). In addition, there is a severe clash of the LigIV catalytic domains as well as other discrepancies when superimposing the SR complex with itself (Extended Data Fig. 9c). The DNA nick identified by rigid-body docking of the LigIV-DNA crystal structure²⁵ is clearly off-center relative to the Ku rings on either side, rendering the Ku with which the LigIV DBD interacts distal to and the other Ku proximal to the nick (Extended Data Fig. 9d). While it has been suggested that NHEJ may only join one strand of a DSB⁴⁵, the positioning of two LigIV molecules with single turnover activity^{46,47} within the SR complex suggests that they are in the correct orientation to sequentially join both strands of the DSB (Extended Data Fig. 9c, Supplementary Video 3).

To validate predictions based on the LR and SR complex structures, *in vitro* ligation assays were performed under conditions used to assemble both synaptic complexes (Methods, Extended Data Fig. 10a). As expected, pre-assembly of DNA-PK prevents ligation when the other NHEJ factors are added in the absence of ATP, presumably because the transition from the LR to the SR complex is blocked. In contrast, ligation by the SR complex is ATP-independent because LigIV is pre-adenylated^{46,47} (Extended Data Fig. 10b,c). Furthermore, joining of both strands occurs more frequently than joining of a single strand even in the absence of ATP and lower amounts of LigIV-XRCC4 (Extended Data Fig. 10b,c), supporting the functional relevance of the two LigIV molecules within the synaptic complexes. Notably, the flexibility of the LigIV catalytic domain⁴⁴ and the position of the aligned DNAs within the SR complex should allow end processing enzymes access since a significant fraction of DSBs repaired by NHEJ require processing prior to ligation^{17,48-50}.

LR to SR synaptic transition

To extrapolate the conformational changes during the transition from LR to SR synapsis, we superimposed the XLF dimers from the two synaptic complexes by excluding DNA-PKcs

(Fig. 4a). While the approximate positions of individual NHEJ factors remain the same relative to each other during the transition, each of the Ku-XRCC4-LigIV subcomplexes rotate to different degrees in order to align the DSB ends with the rotation pivot points in the middle of each XRCC4 CC (Supplementary Video 4). A similar discrepancy is observed when comparing the crystal structures of the human LigIV-XRCC4 complex and its yeast ortholog (Lif1p/Lig4p)³⁴ (Extended Data Fig. 4b), indicating that the bending angle of the intrinsically twistable XRCC4 CC domain can adjust to accommodate different states. Interestingly, the ~16 bp DNA footprint of DNA-PKcs that would be exposed by DNA-PKcs dissociation from the LR complex is precisely the length required to align the two DSB ends in the SR complex (Fig. 4a, Supplementary Video 4), suggesting that the HEAT cradle region of DNA-PKcs serves as a “ruler” by holding the appropriate length of DNA within the DNA-PK sub-complex for subsequent alignment in the SR complex⁵¹. Finally, newly observed interfaces within the SR complex, including XLF-Ku70 and LigIV DBD-Ku70 interfaces, also likely contribute to the transition from the LR complex (Fig. 3c,d).

In summary, we have captured two essential states during NHEJ by high-resolution cryo-EM, allowing us to propose an almost complete reaction cycle of this important DSB repair pathway (Fig. 4b). As described previously, DNA-PKcs is initially recruited by Ku to protect the DNA end from degradation and sequester the DSB for repair by NHEJ^{5,20,21,30,39}. In the LR synaptic state, two DNA-PK complexes engage through the formation of an intricate protein-protein interaction network involving the core NHEJ factors to initially tether the DNA ends. *In trans* autophosphorylation of DNA-PKcs, which triggers their simultaneous dissociation from the DSB ends due to protein-protein or protein-DNA repulsions, serves as a “checkpoint” to ensure that there are two free DNA ends in close proximity. The length of the DNA released from the HEAT cradles of the DNA-PKcs molecules is appropriate for subsequent alignment within the SR complex. Finally, a concerted conformational change is triggered by the restoring force within the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold as well as protein-protein interactions involving XLF-Ku70 and LigIV DBD-Ku70, aligning the DSB ends precisely and allowing the sealing of both strands by the two LigIV molecules on either side of the complex in a tandem manner. Our findings reveal unprecedented insights into DSB repair with implications for modulating the cellular response to radiation and chemotherapy and offer additional opportunities for targeting this and other important PIKK-dependent pathways for cancer therapy.

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Purification of protein factors

DNA-PKcs was purified from the nuclear pellet of HeLa cells⁵², and Ku was purified from baculovirus-infected insect cells⁴⁴ as previously described. The final concentration of DNA-PKcs was 0.5 mg/mL and Ku70/80 1.53 mg/mL. The LigIV-XRCC4 complex was purified

from baculovirus-infected insect cells⁴⁴, and the final concentration was 7.9 mg/mL. XLF was purified as described previously⁵³, and the final concentration was 2 mg/mL.

Assembly of the LR and the SR synaptic complexes

Oligonucleotides used for complex assembly were purchased from Integrated DNA Technology (IDT), and sequences are listed in Supplementary Table 1. Lyophilized oligos were first resuspended in ultra-pure water for a final concentration of 200 μ M. To prepare DNA templates used for complex assembly, complementary strands of each DNA duplex were mixed at a 1:1 molar ratio and a final concentration of 15 μ M in ultra-pure water following the schematic in Extended Data Fig. 1, 7, and 10. They were denatured in boiling water for 5 minutes and cooled down to room temperature for 2 hours. Next, the RNA-biotin bait was annealed with DNA duplex at 45°C for 5 minutes, with a 1.5:1 molar ratio at a final concentration of 2 μ M in ultra-pure water, and then cooled down to room temperature for 45 minutes.

We utilized a duplex DNA substrate with a Y-shape hairpin close to one end to block Ku access^{5,22,30}, and a biotin-labeled RNA-DNA hybrid at the blocked end for pull-down and RNase H elution of DNA-bound complexes⁵⁴ (Extended Data Fig. 1a,d,g and 7a,e,i). To assemble the LR complex with the Y30-T40-c8 DNA-RNA substrate for EM studies, DNA-PK holoenzyme (Ku70/80, DNA-PKcs, dsDNA) was first assembled with two half-DNA substrates separately, then combined to assemble the full complex. Ku70/80 was first incubated with annealed DNA substrate at room temperature for 5 minutes, then diluted in sample buffer (10 mM HEPES pH 7.9, 50 mM KCl, 10 mM MgCl₂, 2.5% glycerol, 1mM DTT, 0.01% NP-40 for negative staining EM and 0.05% NP-40 for cryo-EM. All concentrations are final concentrations after complex assembly). DNA-PKcs was then added, and the sample was incubated for another 5 minutes. Next, two half-complexes were combined and incubated at 37°C for 5 minutes, then cooled down to room temperature in order to help the annealing of the single-strand overhang. LigIV-XRCC4 complex and XLF were then supplied and incubated for 10 minutes. Magnetic streptavidin T1 beads (Invitrogen) were equilibrated with sample buffer and added into the sample for 15 minutes to immobilize the assembled complex. Beads were then washed with sample buffer and eluted for 1 hour with 4 μ L elution buffer (sample buffer with 0.05 unit/ μ L RNase H, New England Biolabs). The SR complexes were assembled using a similar protocol, with adjusted protein concentrations and additional incubation of combined half-complexes with single-strand complement DNA to assemble the holo-complex when necessary.

The negative staining samples of the LR complex contained ~700 nM Ku70/80 and DNA-PKcs, ~2.1 μ M LigIV-XRCC4, ~500 nM XLF, as well as ~350 nM of each half-DNA substrate. The SR complex negative staining samples contained ~700 nM Ku70/80, ~1 μ M LigIV-XRCC4 and XLF, as well as ~350 nM of each half-DNA substrate. The first dataset of the LR complex was collected on the carbon-coated C-flat grids (Electron Microscopy Sciences). The rest of the cryo-EM data collections were done using Graphene-Oxide (GO) grid⁵⁵ to improve the signal-to-noise ratio.

Electron microscopy

400 mesh copper grid (Electron Microscopy Sciences) with continuous carbon supporting layer was used to prepare negative stain samples. Before sample deposition, the grid was plasma-cleaned for 10 seconds using a Solarus plasma cleaner (Gatan) equipped with air at 25W power. Purified complex sample (3 μ L) was cross-linked for 5 minutes with 0.05% glutaraldehyde and then loaded onto the grid. The grid was incubated in a 100% humidity chamber for 10 minutes to absorb the sample particles. After the incubation, the grid was stained by four successive 35 μ L drops of 2% (w/v) uranyl formate solution for 5s, 10s, 15s, and 20s, then blotted till dryness. Data was collected using a JEOL 1400 transmission electron microscope operating at 120 kV at a nominal magnification of $\times 30,000$ (3.71 $\text{\AA}$ per pixel). Leginon data collection software⁵⁶ was used for data collection on a Gatan 4k $\times$ 4k CCD camera using low-dose procedures (20 e⁻ $\text{\AA}^{-2}$ exposures). The defocus range was -1.5 to -3 μ m.

The same cross-linking procedure was used for cryo-EM sample preparation (total volume 4 μ L). For the first dataset of the LR complex that was collected on C-flat grids, the grid was prepared ahead of time by floating thin carbon film onto the holey carbon grid (CF-3.5/1-4C, C-flat grid with 3.5 μ m hole and 1 μ m space, Electron Microscopy Sciences) and was plasma cleaned for 10s at 5 W power using the Solarus plasma cleaner (Gatan) right before sample loading. Cross-linked sample was loaded onto the grid and incubated in the 4°C, 100% humidity chamber of Vitrobot (FEI) under low illumination conditions. Grid was blotted for 4 seconds at 25 force, plunge-frozen in liquid ethane, and stored in liquid nitrogen. For the other datasets that were collected on GO grids, a thin layer of GO was applied onto either the same C-flat grids or 200 mesh Quantifoil holey carbon grids (3.5/1 copper and 2/1 copper/gold, Electron Microscopy Sciences, all yielding similar results) and the grids are prepared right before sample loading.

For all four datasets collected on the LR complex, JEM-3200FS Field Emission Electron Microscope (JEOL) operating at 200 kV was used with K2 Summit direct electron detector (Gatan) at a nominal magnification of $\times 30000$ (pixel size 1.1 $\text{\AA}$, counting mode). The defocus range was -2 to -4 μ m. MSI-Raster2 application of the Leginon data collection software⁵⁶ was used for collecting 17,114 movie series in total. 30-frame exposures were taken at 0.3 s per frame (9 s total exposure time), using a dose rate of 8 e⁻ per pixel per second, corresponding to a total dose of 76.5e⁻ $\text{\AA}^{-2}$ per movie series.

Two datasets were collected (32,723 movie series in total) for the SR complex at The Pacific Northwest Center for Cryo-EM (PNCC) using Serial EM⁵⁷. Titan Krios-3 TEM (Thermo Fisher) operating at 300 kV was used with K3 direct electron detector (Gatan) at $\times 30000$ magnification (0.83 $\text{\AA}$ /pixel, counting mode, 0.415 $\text{\AA}$ /pixel, super-resolution mode). The defocus range was -1.5 to -4 μ m. 50-frame exposures were taken during 2.13s total exposure time, using a dose rate of 15 e⁻ per pixel per second, corresponding to a total dose of $\sim 46\text{e}^- \text{\AA}^{-2}$ per movie series.

Image processing and three-dimensional reconstruction

Negative stain data pre-processing was performed using the Appion processing environment⁵⁸. Automatic particle selection was performed on the micrographs with the Difference of Gaussians (DoG) particle picker⁵⁹. CTFFind4⁶⁰ was used to estimate the contrast transfer function (CTF) of each micrograph and flip the phases. For the LR complex, 36,819 particles were extracted and stacked with a box size of 144×144 pixels ($3.71 \text{ \AA}/\text{pix}$). Two-dimensional classification was conducted using iterative multivariate statistical analysis and multi-reference alignment analysis (MSA-MRA) within the IMAGIC software⁶¹. Three-dimensional (3D) reconstruction of negative stained data was performed using an iterative multi-reference projection-matching approach containing libraries from the EMAN2 software package⁶². The initial 3D model was generated using cryoSPARC⁶³. A similar protocol was used to generate the SR complex initial model. 60,194 particles were extracted and stacked with a box size of 128×128 pixels.

RELION 3.1 was then used for all the pre-processing, 3D classification, model refinement, post-process, and local-resolution estimation jobs⁶⁴. To pre-process the cryo-EM data, all the movie frames were aligned using RELION's own implementation. After motion correction, particles were automatically picked using Gautomatch (developed by K. Zhang, MRC Laboratory of Molecular Biology, Cambridge, UK), and the local CTF of each micrograph is determined using Gctf⁶⁵ or CTFFIND-4.1⁶⁶.

For the 3D reconstruction of the LR complex, 1,119,381 particles were automatically picked and supplied to an initial round of 3D classification step with the negative stain reconstruction map (low-pass filtered to $30 \text{ \AA}$) as an initial reference. Larger angular sampling interval, offset search range (pix), and offset search step (pix) were used for the first 50 iterations (15° , 10° , and 2° respectively), then followed by 190 iterations of default classification (7.5° , 10° , and 1°). Class 4 (247,254 particles) showed sharp structural features, so it was selected for further processing. Class 3 (380,064 particles) was the second-best class and was subjected to two rounds of reference-free 2D classification (20 classes, 50 iterations) to further remove bad particles. Six classes that showed sharp features and two copies of DNA-PKcs in the 2D class averages (82,530 particles) were selected and combined with class 4. Next, the selected particles were 3D auto-refined (with C2 symmetry), re-centered, and re-extracted without binning ($1.1 \text{ \AA}$ per pixel, box size 432). Another round of 3D auto-refinement on this stack was performed with a soft mask applied around the whole complex, resulting in a $5.54 \text{ \AA}$ resolution reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion⁶⁷. Particles were then subjected to per-particle CTF refinement (per-particle defocus and per-micrograph astigmatism estimation, as well as beam tilt estimation), followed by Bayesian particle polishing⁶⁸. 3D auto-refinement against the shiny particles yielded a $4.85 \text{ \AA}$ resolution map, which was further improved to $4.6 \text{ \AA}$ by an additional round of CTF refinement (estimating both defocus, astigmatism, and magnification anisotropy) and Bayesian particle polishing. This map was used as the overall map for deposition.

The overall map of the LR complex contains two copies of a relatively rigid DNA-PK complex (DNA-PKcs, Ku70/80 and DNA) and a very flexible density corresponding to XRCC4-LigIV-XLF. To focus on the body of the rigid region (DNA-PKcs, Ku70/80, DNA and the first BRCT domain of LigIV) in the complex, a partial soft mask was applied to one copy of the corresponding volume, and C2 symmetry expansion was applied to align two copies of the rigid body into one dataset. Signal subtraction with the mask results in a 660,050-particle stack that contained all copies of rigid bodies. Next, the particles were binned by 2 again (2.2 Å per pixel, box size 216) and 3D auto-refined to 4.40 Å resolution. All of the 3D auto-refine steps after symmetry expansion were performed locally with an initial angular sampling interval of 3.7°. Subsequent two-class 3D classification with no alignment was performed, and the class with better features and higher resolution (333,166 particles) was selected to be unbinned (1.1 Å per pixel, box size 432) auto-refined and post-processed to 4.1 Å. Local resolution of the map was also estimated within RELION 3.1⁶⁹.

In order to generate a focused signal-subtracted map for the flexible body (XRCC4-LigIV BRCT-XLF), a similar strategy was applied to the same polished stack without applying symmetry expansion. The binned stack was then subjected to 7-class 3D classification without applying any alignment. The best class (43,510 particles) was further refined and post-processed to 9.3 Å.

A similar strategy was used for the processing of the SR complex dataset. However, due to the limited resolution obtained, even after the final steps, all of the particles were binned by 2 (1.66 Å/pxl) during the whole process for better alignment without further un-binning. 1,766,936 particles were picked for direct 3D classification, and 175,866 particles were selected and refined to 8.4 Å after CTF refinement and polishing. This map was used as the complex map for deposition. Next, the complex was divided into three bodies for further focused refinement: 1. A C2 body similar to the LR complex low-resolution region (XRCC4-LigIV BRCT-XLF); 2. A C1 body containing distal Ku70/80, LigIV catalytic domain and their DNA footprints; 3. A C1 body containing proximal Ku70/80 and its DNA footprints. Signal subtraction and further classification and refinement improved the resolution of all three bodies. Specifically, body 1 was refined to 7.1 Å (45,934 particles), body 2 6.8 Å (57,170 particles) and body 3 6.9 Å (54,793 particles)

UCSF Chimera was used for all the volume segmentation, figure and movie generation, and automatic rigid-body docking processes⁷⁰. In parallel with post-processing done in RELION3.1, DeepEMhancer was applied on the refined maps as well to better correct local B-factor and yielded cleaner maps for model building and docking⁷¹.

Model Building

Sequence alignment of all of the subunits (DNA-PKcs, Ku70, Ku80, XLF, XRCC4, and LigIV) was performed using CLC Sequence Viewer 7 (Supplementary Figs. 1-6). To build the structure model of the rigid region in the LR complex (DNA-PKcs, Ku70/80, DNA and 1st BRCT domain of LigIV), the published crystal and EM structures of DNA-PKcs (PDB 5LUQ²⁰, 6ZHA³⁰), Ku70/80 complex bound with XLF KBM (PDB 6ERH²⁹, without DNA) and the N-terminal BRCT domain of LigIV in complex with XRCC4 (PDB 3II6²⁴) were

referenced as homology models. The DNA was first built as standard B-form DNA using Coot⁷², then fitted into the density using the interactive molecular-dynamics flexible fitting software ISOLDE⁷³. The DNA sequences were arbitrarily assigned as the poly-T overhang end of input double-strand DNA. Next, the models were manually inspected and refined to fit the model using both Coot and ISOLDE. The published cryo-EM structure model of DNA-PKcs (PDB 5W1R²¹) and DNA-PK holoenzyme (PDB 5Y3R⁵, 7K0Y³⁹), as well as the crystal structure model of DNA-bound Ku70/80 complex (PDB 1JEY²²), were also referenced throughout the process^{74,75}. Residues 810-845, 2577-2605, and 2721-2773 in DNA-PKcs have not been modeled in any of the published structures to date and were modeled *de novo* using Coot and ISOLDE. The registers can be clearly identified as most of the densities and bulky side-chains aligned well between the map and the model. To further validate our model, the deep-learning based multiple sequence alignment module that is built-in trRosetta⁷⁶ was used against the corresponding residues in DNA-PKcs to predict the secondary structure *de novo* and the resulted location of helices and loops aligned well with our model. Meanwhile, residues 585-601, 2902-2914, and 4008-4037 that were resolved in the previous DNA-PKcs models lacked corresponding densities in the LR complex density and were removed from the final reconstruction. DNA-PKcs kinase domain was superimposed with the structure of mTOR kinase binding to ADP (PDB 4JSV⁴⁰). The location of the ADP molecule aligned well with the LR complex density, so it was directly used to dock in the map. Due to the resolution limitation, the register of residues was hard to determine unambiguously within part of the DNA-PKcs cradle domain, the DNA-PKcs FRB region, the BRCT domain, and the Ku80 vWA domain that binds to XLF peptide. For those regions, only the secondary motifs are flexibly fitted into density with minimal disruption of existing model constraints. The density that could not be incorporated into the structure of DNA-PKcs is located in the vicinity of the FAT domain (α -helix 2965-2977, 2991-3005, and 3008-3018) and is believed to be part of unstructured XRCC4 C-terminal tail. Based on the distribution of bulky side chains and XRCC4 sequence conservation, residue 270-278 is assigned to this extra density and manually built using Coot and ISOLDE (Supplementary Fig. 5). Manual refinement and the Phenix real space refinement⁷⁷ were performed iteratively to further refine the model.

To build the model for the flexible body in the LR complex (XRCC4, XLF, and the 2nd BRCT domain of LigIV), the crystal structure of XLF-XRCC4 filament (PDB 3Q4F³², with one copy of XLF dimer and two copies of XRCC4 dimer) were rigid-body docked in using UCSF Chimera as a template. Next, crystal structures of XLF (PDB 2R9A²³) and XRCC4 in complex with 2nd BRCT domain of LigIV (PDB 3II6²⁴) were aligned to the filament template, respectively. With this strategy, the interaction surfaces between XRCC4 and XLF were well-maintained, but the coiled-coil of XRCC4 and the BRCT domain of LigIV was off a little compared to the density. To better fit this part into the relative low-resolution map without disrupting the overall model architecture, ISOLDE was used to apply adaptive distance restraints between nearby atoms within this part of the model, and the helices and BRCT domains were then flexibly fitted into the density.

The complete model of the LR complex is composed of two copies of the rigid body and one copy of the flexible body. To generate such models, the above models were rigid-body docked into the composed map using UCSF Chimera, and the loop between two BRCT

domains of LigIV was used to guide the linkage of the two domains, which were originally fitted into two bodies separately. The resolution of the linker loop was not good enough to perform *de novo* model building, but the terminal residues of the two BRCT domains were located in close proximity in the composed model. Coot was used to manually build the one missing residue (TYR765) and link them with minimal extra disruption.

Subunits from the LR complex (no DNA) were used to flexibly fit in the density of the SR complex model. The catalytic domain model of LigIV (PDB: 6BKF²⁵) was docked into the corresponding map region as well. Next, ISOLDE was used to flexibly refine the models with adaptive distance restraints. An intact B-form DNA was mutated into the substrate sequence and then used to flexibly fit for the entire DNA density using ISOLDE. After that, the corresponding nicks and nucleotide deletions were introduced to match the dsDNA exactly with the designed substrate. The position of the nick aligned well with the LigIV catalytic domain binding site, and the flexible single strand -TT- linkers between Ku70/80 and LigIV catalytic domain were also aligned with the worst density region that corresponds to dsDNA. For clarification, all subunits that are located on the left in Fig. 1 and Fig. 3 are named copy A, and subunits on the right are named copy B.

***In vitro* Ligation Assay**

The sequences of Cy5-labeled DNA substrates, purchased from Integrated DNA Technology (IDT), are listed in Supplementary Table 1. DNA substrates were annealed as described above. The substrate architecture and different experimental conditions are shown as in Extended Data Fig. 10.

The LR and SR complexes with Cy5-labeled substrate were assembled similarly as described for EM sample preparation above, while buffer compositions before and after pull-down were modified for the reaction. Specifically, buffer with no MgCl₂ and 50 mM EDTA was used during complex assembly in order to prevent non-specific LigIV ligation. After the intact complex was assembled and bound to the streptavidin-coated magnetic beads, samples were washed with buffer containing 5 mM MgCl₂ without EDTA. 1 mM ATP was then added to the sample, and the reaction was carried out for 10 minutes at room temperature. For the control lanes, the complex assembly buffer remained unmodified, but various buffer conditions used after bead pull-down were modified accordingly. Next, the beads were boiled at 95°C for 5 minutes to elute off the DNA, and Proteinase K was added to allow incubation at 37°C for 15 minutes. 50% volume of formamide and 10 mM of EDTA were added to prepare the sample for denatured urea gel (TBE gel with 8% acrylamide (19:1)). Gel electrophoresis was performed at 250V for 40 minutes and scanned using a Sapphire Biomolecular Imager (Azure) at optimal absorbance.

Extended Data

Extended Data Fig. 1. Optimization of the LR synaptic complex assembly with various DNA substrates.

a, Schematic showing the Y-35 blunt-end DNA substrate. Complex assembly was attempted (supplying DNA-PKcs, Ku, XLF, and LigIV-XRCC4) prior to purification via RNase-H elution. **b**, A representative negative staining raw micrograph of the complex assembled as described in **a**. The raw micrograph is representative of 24 micrographs. **c**, Representative 2D class-averages of the complex assembled as described in **a**, showing the appearance of only DNA-PK complex despite the addition of XLF and LigIV-XRCC4. **d-f**, Same procedure as **a-c**, showing the complex assembly with the same Y-35 substrate, but adding XLF and LigIV-XRCC4 to purified DNA-PK complex following RNase-H elution. The raw micrograph is representative of 27 micrographs. In **f**, 2D class-averages representing the characteristic view of scarce but existing LR complex are obtained. **g-i**, Same procedure as **a-c**, showing the complex assembly using Y30-T40-c8 DNA substrate with 40 nt flexible poly-T and 8 bp of complementary ends as 3' overhang. While the single-stranded poly-T overhang and the 8-bp complementary region contribute to complex stability, they are not

observed in any part of the reconstructed density map, presumably because these ssDNA tethers are too flexible to be aligned with the rest of the complex. The raw micrograph is representative of 24 micrographs. The complex was assembled prior to RNase-H elution as described in **a**. In **i**, the majority of the 2D classes correspond to the LR complex. **j**, A representative cryo-EM raw micrograph (out of 17,114 in total) of the LR complex assembled with the Y30-T40-c8 DNA substrate shown in **g**. **k**, Representative 2D class averages of particles (329,784 in total) contributing to the final reconstruction of the LR complex. **l**, Silver-stained SDS-PAGE (4–12% gradient, biologically replicated three times) showing the input purified subunits (Ku, DNA-PKcs, LigIV-XRCC4, and XLF) and the RNase-H purified LR and SR complex for cryo-EM data collection. All representative micrographs in **b**, **e**, **h**, **j** are from at least three biologically replicated experiments. For gel source data, see Supplementary Figure 1. **m**, Protein-protein interaction network between the components of the LR complex. Major unmodeled regions are shown in gray. Well-documented hetero- or homo-dimers are grouped by red dashed lines. Alternative protein-protein interactions are depicted by black dashed lines. The globular domain within the Ku80 C-terminal region (CTR) is completely flexible in the LR complex, and we do not see evidence of the Ku80 CTR domain swap observed by Chaplin *et al.*³⁰. The putative distance between one Ku80 CTR globular domain and the other copy of the Ku80 C-terminal helix is too far to be reached by the 18-amino acid linker within Ku80. Abbreviations: N-HEAT: N-terminal HEAT domain; M-HEAT: middle HEAT domain; KD: kinase domain; vWA: von Willebrand A domain; CTD: C-terminal domain; DBD: DNA-binding domain; NTD: N-terminal domain; OBD: OB-fold domain; BRCTs: tandem BRCT domains; HD: head domain; CC: Coiled-coil domain.

Extended Data Fig. 2. Data-processing scheme of the LR synaptic complex sample.

a, Flow chart of the cryo-EM data processing procedure. The gold-standard Fourier Shell Correlation (FSC) curves (0.143 cutoff) show the final resolution of the holo-complex and each body. **b**, sample maps and fitted models of DNA-PKcs (olive) and dsDNA substrate (cyan) from the LR complex are shown at 4.1 Å resolution. Maps are shown as transparent surfaces, and models are shown as sticks, respectively.

Extended Data Fig. 3. Comparing the structure of Ku among the LR synaptic complex, the SR synaptic complex, and previously published models.

a, Ku structure in the LR complex showing outward rotations of both Ku70 and Ku80 vWA domains. **b**, Crystal structure of XLF KBM bound Ku showing the outward rotation of only Ku80 vWA domain²⁹. **c**, Crystal structure of apo Ku showing no rotation of either Ku70 or Ku80 vWA domains²². **d**, Conformation of Ku shown in the cryo-EM structure of apo DNA-PK complex³⁰. Ku70 vWA domain is rotated outward, triggered by binding of DNA-PKcs. **e-f**, Two copies of XLF KBM bound Ku in the SR complex. The conformation of both copies is the same as the one in the LR complex (**a**), despite the fact that DNA-PKcs is not present. Color codes for Ku70 and Ku80 are the same as in Fig. 1.

Extended Data Fig. 4. Comparing the structure of LigIV-XRCC4-XLF scaffold among the LR synaptic complex, the SR synaptic complex, and previously published models.

a, Structure of XRCC4-XLF from the LR complex is shown in comparison with XRCC4-XLF filamentous repeat crystal structures and the ones from the SR complex (both copies). The XLF dimer is used to align all of the models shown here. Solid lines are aligned with the coiled-coils (CC) of XLF (vertical) and XRCC4 (tilted), and the angles in between are shown, respectively. Dashed lines are aligned with the C-terminal half of CC of XRCC4 when full helices are present, and the bending angles are shown as well. **b**, XRCC4 in the crystal structure of human and yeast LigIV-XRCC4 complex^{24,34} are shown after aligning with XRCC4 in the LR complex shown in **a**. The bending of XRCC4 CC is more similar to the one in the SR than in the LR complex. **c**, Structure of LigIV-XRCC4 complex from the LR complex is shown in comparison with human and yeast LigIV-XRCC4 crystal structures and ones from the SR complex (both copies). Color codes for XLF, XRCC4, and LigIV BRCT domains are the same as in Fig. 1.

Extended Data Fig. 5. Surface electrostatic potential and conservation of different areas in the LR synaptic complex.

a, Close-up view of the interaction surface between LigIV N-BRCT domain and Ku70 vWA domains colored by sequence conservation. **b**, Close-up view of the DNA-PKcs-DNA-PKcs interaction surface colored by sequence conservation. **c**, Surface electrostatic potential view of DNA-PKcs near its FAT domain, showing its negatively charged interface between XRCC4 C-terminal region (ribbon). The approximate path of the XRCC4 C-terminal peptide containing multiple phosphorylation sites is depicted. The sphere depicts the location of a cancer-associated truncation mutation that occurs at the interface. **d**, Surface electrostatic potential view of DNA-PKcs DEB and DEB-A helix. The DNA-interaction surface is positively charged. When models are not colored by either surface electrostatic potential or sequence conservation, the color codes are the same as in Fig. 1. We cannot rule out the unlikely possibility that the stabilization of the DEB helix is due to the presence of a 3' overhang that existed in our DNA substrate design (Extended Data Fig. 1g).

Extended Data Fig. 6. Comparing the dimerization of DNA-PKcs in the LR synaptic complex with other PIKK family dimers.

a, Structure of the two DNA-PKcs in the LR complex. The Kinase domain (KD) is aligned with the homologous domains in **b** and **c** as an anchor point. **b-c**, Dimer of ATR-ATRIP (**b**) and ATM (**c**) showing aligned KD and corresponding N-terminus HEAT regions in the aligned copy. The symmetric-look front views are shown at the bottom left corner. Each protomer of ATR-ATRIP and ATM and colored the same as the corresponding DNA-PKcs protomer, in olive (the aligned copy) and dark khaki (the other copy). **d**, Domain organization of DNA-PKcs compared with ATR and ATM. Abbreviations are the same as in Fig. 1. In our model, both the ABCDE (T2609, S2612, T2620, S2624, T2638, and T2647) and the PQR (S2023, S2029, S2041, S2053, and S2056) phosphorylation sites are located within disordered loops of DNA-PKcs 2606-2720 and 1993-2084, respectively (Fig. 2b,d). The kinase active center from the opposite side cannot reach most of the ABCDE sites unless the YRPD-Interaction (YRPD-I) loop (residue 2586-2604) is peeled off from the YRPD motif (Fig. 2d). In turn, this conformational change potentially disrupts the DNA-PKcs-DNA-PKcs dimerization interface through loop 2569-2585 (Fig. 1d). Similarly, some PQR sites are located too far from the trans kinase active center. PQR autophosphorylation induced changes could have a direct impact on the Ku80 CTR-DNA-PKcs interface at the bottom of the cradle (Fig. 2d), potentially inducing the domain-swap of Ku80 observed by Chaplin *et al.*³⁰.

Extended Data Fig. 7. Optimization of the SR synaptic complex assembly with various DNA substrates.

a, Schematic showing the Y30-c4 DNA substrate with 4nt 3' complementary overhang. The complex was assembled prior to RNase-H elution. **b**, A representative negative staining raw micrograph of the complex assembled as described in **a**. The raw micrograph is representative of 23 micrographs. **c**, Representative 2D class-averages of the complex assembled as described in **a**. **d**, Cryo-EM map reconstructed from a small dataset using Y30-c4 DNA substrate shown in **a**. The map is colored by local resolution estimation. **e-h**, Same procedure as **a-d**, showing the complex assembly with the Y30 blunt end substrate. Stably assembled SR complexes on DNA substrates with either complementary or blunt ends indicates that these complexes are stable in the absence of any bridging effect from DNA. The raw micrograph is representative of 24 micrographs. **i-l**, Same procedure as **a-d**, showing the complex assembly with the Y14-T2-c20-n10-10 substrate, with one central single non-ligatable nick. The raw micrograph is representative of 24 micrographs. Strand e is added at last after mixing the two halves together with NHEJ factors. **m**, A representative cryo-EM raw micrograph (out of 32,723 total images) of the SR complex assembled with

the Y14-T2-c20-n10-10 DNA substrate shown in **i**. **n**, Representative 2D class averages of particles (175,866 in total) contributing to the final reconstruction of the SR complex. All representative micrographs in **b**, **f**, **j**, **m** are from at least three biologically replicated experiments. **o**, Protein-protein interaction network between the components of the SR complex. Major unmodeled regions are shown in gray. Well-documented hetero- or homo-dimers are grouped by red dashed circles. Alternative protein-protein interactions are depicted by black dashed lines.

Extended Data Fig. 8. Data-processing scheme of the SR synaptic complex sample.

Flow chart of the cryo-EM data processing procedure. The gold-standard FSC curves (0.143 cutoff) show the final resolution of the holo-complex and each body.

Extended Data Fig. 9. Surface conservation of different areas in the SR synaptic complex.

a, Close-up view of the interaction surface between LigIV DBD and Ku70 vWA domain colored by sequence conservation. DNA-PKcs clashes with LigIV DBD when Ku is aligned between the LR and SR complex **b**, Close-up view of XLF CC at its C-terminal tip colored by sequence conservation. When models are not colored by sequence conservation, the color codes are the same as in Fig. 3. **c**, Superimposition of two asymmetric SR complexes after a 180° flip shown in front (top) and top (bottom) views. XLF homodimer is used for aligning the two conformers. LigIV catalytic domains are hidden for clarity purposes. The transition from the apo state to the flipped state indicates potential conformational changes during the tandem ligation. Paths of DNA are also highlighted by dashed lines. **d**, Close-up view showing the relative positions of the two off-centered nicks between the two conformers. The two preferential nick positions are separated by approximately 4 bp. Consistently, dsDNA with 4nt 3' overhang, a major end-processing product of the NHEJ nuclease–Artemis⁷⁸, is reported to be a favored substrate for NHEJ⁵⁰. Intriguingly, our model suggests that dsDNA with a 4 nt 3' overhang will experience minimum DNA translocation to accommodate the two ligation steps (Supplementary Video 3).

Extended Data Fig. 10. Both LR and SR synaptic complexes are able to perform double ligation during NHEJ *in vitro*.

a, Substrate design for the ligation assay. An Internal Cy5 label is added to only the right half of the substrate to visualize the ligation products. A 4 nt 3' complementary overhang has been introduced on both sides of the substrate. **b**, Denaturing gel analysis of end-joining by the LR complex. 100nM of DNA, 200 nM of DNA-PKcs, and Ku70/80, 500nM of XLF, and 70nM of X4L4 were added, respectively. Asterisk indicates alternative secondary structure or impurity of the cy5-labeled oligo. Size of the DNA substrates and ligation products are labeled on the left (unit: bp). **c**, Denaturing gel analysis of end-joining by the SR complex. The final factor concentrations are the same as in **b**. For gel source data, see Supplementary Figure 1. Similar conditions for either of the gel have been replicated as biological replicates for two times.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Jason Pattie for computer support, Janette Meyers, Rose Marie Haynes, and Harry Scott at the PNCC for data collection support, and Dale Ramsden for the generous gift of a rabbit anti-phosphoT2609 reagent. We are grateful to Amy Rosenzweig and Ishwar Radhakrishnan for helpful discussion and comments on the manuscript. This work was supported by a Cornew Innovation Award from the Chemistry of Life Processes Institute at Northwestern University (to Y He), a Catalyst Award by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to Y He), an Institutional Research Grant from the American Cancer Society (IRG-15-173-21 to Y He), an H Foundation Core Facility Pilot Project Award (to Y He), a Pilot Project Award under U54CA193419 (to Y He), and NIH grant R01 GM135651 (to Y He). S Chen is supported by the Molecular Biophysics Training Program from NIGMS/NIH (5T32 GM008382). A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental

Research. This work used the Sapphire imager from the Northwestern University Keck Biophysics Facility funded by NIH grant 1S10OD026963-01, as well as the resources of the Northwestern University Structural Biology Facility, which is generously supported by NCI CCSG P30 CA060553 grant awarded to the Robert H. Lurie Comprehensive Cancer Center. The Gatan K2 direct electron detector was purchased with funds provided by the Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust. Work in the Lees-Miller lab was supported by Canadian Institutes of Health grant #16939 and the Engineered Air Chair in Cancer Research. Work in the Tomkinson lab was supported by NIH grant R01GM047251, as well as the University of New Mexico Comprehensive Cancer Center supported by NCI CCSG P30 CA118100 grant. The collaboration between the He, Lees-Miller, and Tomkinson labs was supported by NCI P01 CA092584.

Data availability

Cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-23510 (overall LR complex), EMD-23511 (DNA-PK-N-BRCT in the LR complex), EMD-23512 (LigIV-XRCC4-XLF-XRCC4-LigIV in the LR complex), EMD-23509 (overall SR complex), EMD-23513 (Distal Ku-LigIV Catalytic domain in the SR complex), EMD-23514 (Proximal Ku in the SR complex), and EMD-23515 (LigIV-XRCC4-XLF-XRCC4-LigIV in the SR complex). Model coordinates have been deposited in the Protein Data Bank under accession numbers 7LT3 (the LR complex) and 7LSY (the SR complex).

References

1. Aplan PD Causes of oncogenic chromosomal translocation. *Trends Genet.* 22, 46–55 (2006). [PubMed: 16257470]
2. Zhao B, Rothenberg E, Ramsden DA & Lieber MR The molecular basis and disease relevance of non-homologous DNA end joining. *Nat. Rev. Mol. Cell Biol* 32, 66–17 (2020).
3. Malu S, Malshetty V, Francis D & Cortes P Role of non-homologous end joining in V(D)J recombination. *Immunol. Res* 54, 233–246 (2012). [PubMed: 22569912]
4. Gottlieb TM & Jackson SP The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 72, 131–142 (1993). [PubMed: 8422676]
5. Yin X, Liu M, Tian Y, Wang J & Xu Y Cryo-EM structure of human DNA-PK holoenzyme. *Cell Res.* 87, 237 (2017).
6. Mari P-O et al. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci. U.S.A* 103, 18597–18602 (2006). [PubMed: 17124166]
7. Yano K-I & Chen DJ Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. *Cell Cycle* 7, 1321–1325 (2008). [PubMed: 18418068]
8. DeFazio LG, Stansel RM, Griffith JD & Chu G Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.* 21, 3192–3200 (2002). [PubMed: 12065431]
9. Chen L, Trujillo K, Sung P & Tomkinson AE Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J. Biol. Chem* 275, 26196–26205 (2000). [PubMed: 10854421]
10. Ramsden DA & Gellert M Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *EMBO J.* 17, 609–614 (1998). [PubMed: 9430651]
11. Riballo E et al. XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucl. Acids Res* 37, 482–492 (2009). [PubMed: 19056826]
12. Andres SN et al. A human XRCC4-XLF complex bridges DNA. *Nucl. Acids Res* 40, 1868–1878 (2012). [PubMed: 22287571]
13. Lees-Miller SP, Chen YR & Anderson CW Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T antigen, mouse p53, and the human Ku autoantigen. *Molecular and Cellular Biology* 10, 6472–6481 (1990). [PubMed: 2247067]

14. Meek K, Douglas P, Cui X, Ding Q & Lees-Miller SP trans Autophosphorylation at DNA-Dependent Protein Kinase's Two Major Autophosphorylation Site Clusters Facilitates End Processing but Not End Joining. *Molecular and Cellular Biology* 27, 3881–3890 (2007). [PubMed: 17353268]
15. Uematsu N et al. Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol* 177, 219–229 (2007). [PubMed: 17438073]
16. Mahaney BL, Hammel M, Meek K, Tainer JA & Lees-Miller SP XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochem. Cell Biol* 91, 31–41 (2013). [PubMed: 23442139]
17. Chang HHY et al. Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J. Biol. Chem* 291, 24377–24389 (2016). [PubMed: 27703001]
18. Conlin MP et al. DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining. *CellReports* 20, 2810–2819 (2017).
19. Zhao B et al. The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. *Nature Communications* 10, 3588–12 (2019).
20. Sibanda BL, Chirgadze DY, Ascher DB & Blundell TL DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science* 355, 520–524 (2017). [PubMed: 28154079]
21. Sharif H et al. Cryo-EM structure of the DNA-PK holoenzyme. *Proc. Natl. Acad. Sci. U.S.A* (2017). doi:10.1073/pnas.1707386114
22. Walker JR, Corpina RA & Goldberg J Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412, 607–614 (2001). [PubMed: 11493912]
23. Andres SN, Modesti M, Tsai CJ, Chu G & Junop MS Crystal Structure of Human XLF: A Twist in Nonhomologous DNA End-Joining. *Mol. Cell* 28, 1093–1101 (2007). [PubMed: 18158905]
24. Wu P-Y et al. Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Molecular and Cellular Biology* 29, 3163–3172 (2009). [PubMed: 19332554]
25. Kaminski AM et al. Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nature Communications* 9, 2642 (2018).
26. Graham TGW, Walter JC & Loparo JJ Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol. Cell* 61, 850–858 (2016). [PubMed: 26990988]
27. Graham TGW, Carney SM, Walter JC & Loparo JJ A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nat Struct Mol Biol* 412, 607 (2018).
28. Carney SM et al. XLF acts as a flexible connector during non-homologous end joining. *eLife* 9, (2020).
29. Nemoz C et al. XLF and APLF bind Ku80 at two remote sites to ensure DNA repair by non-homologous end joining. *Nat Struct Mol Biol* 25, 971–980 (2018). [PubMed: 30291363]
30. Chaplin AK et al. Dimers of DNA-PK create a stage for DNA double-strand break repair. *Nat Struct Mol Biol* 592, 145–7 (2020).
31. Andres SN & Junop MS Crystallization and preliminary X-ray diffraction analysis of the human XRCC4-XLF complex. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67, 1399–1402 (2011).
32. Ropars V et al. Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc. Natl. Acad. Sci. U.S.A* 108, 12663–12668 (2011). [PubMed: 21768349]
33. Hammel M et al. XRCC4 Protein Interactions with XRCC4-like Factor (XLF) Create an Extended Grooved Scaffold for DNA Ligation and Double Strand Break Repair. *J. Biol. Chem* 286, 32638–32650 (2011). [PubMed: 21775435]
34. Doré AS et al. Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair (Amst)* 5, 362–368 (2006). [PubMed: 16388993]
35. Costantini S, Woodbine L, Andreoli L, Jeggo PA & Vindigni A Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)* 6, 712–722 (2007). [PubMed: 17241822]

36. Normanno D et al. Mutational phospho-mimicry reveals a regulatory role for the XRCC4 and XLF C-terminal tails in modulating DNA bridging during classical non-homologous end joining. *eLife* 6, 1093 (2017).
37. Tate JG et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucl. Acids Res* 47, D941–D947 (2019). [PubMed: 30371878]
38. Lees-Miller JP et al. Uncovering DNA-PKcs ancient phylogeny, unique sequence motifs and insights for human disease. *Prog. Biophys. Mol. Biol* (2020). doi:10.1016/j.pbiomolbio.2020.09.010
39. Chen X et al. Structure of an activated DNA-PK and its implications for NHEJ. *Mol. Cell* (2020). doi:10.1016/j.molcel.2020.12.015
40. Yang H et al. mTOR kinase structure, mechanism and regulation. *Nature* 497, 217–223 (2013). [PubMed: 23636326]
41. Bao ZQ, Jacobsen DM & Young MA Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis. *Structure* 19, 675–690 (2011). [PubMed: 21565702]
42. Bareti D et al. Structures of closed and open conformations of dimeric human ATM. *Sci Adv* 3, e1700933 (2017). [PubMed: 28508083]
43. Rao Q et al. Cryo-EM structure of human ATR-ATRIP complex. *Cell Res.* 28, 143–156 (2018). [PubMed: 29271416]
44. Hammel M et al. An Intrinsically Disordered APLF Links Ku, DNA-PKcs, and XRCC4-DNA Ligase IV in an Extended Flexible Non-homologous End Joining Complex. *J. Biol. Chem* 291, 26987–27006 (2016). [PubMed: 27875301]
45. Pryor JM et al. Ribonucleotide incorporation enables repair of chromosome breaks by nonhomologous end joining. *Science* 361, 1126–1129 (2018). [PubMed: 30213916]
46. Riballo E et al. Cellular and biochemical impact of a mutation in DNA ligase IV conferring clinical radiosensitivity. *J. Biol. Chem* 276, 31124–31132 (2001). [PubMed: 11349135]
47. Wang Y, Lamarche BJ & Tsai M-D Human DNA ligase IV and the ligase IV/XRCC4 complex: analysis of nick ligation fidelity. *Biochemistry* 46, 4962–4976 (2007). [PubMed: 17407264]
48. Tsai CJ, Kim SA & Chu G Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc. Natl. Acad. Sci. U.S.A* 104, 7851–7856 (2007). [PubMed: 17470781]
49. Gerodimos CA, Chang HHY, Watanabe G & Lieber MR Effects of DNA end configuration on XRCC4:DNA ligase IV and its stimulation of Artemis activity. *J. Biol. Chem* (2017). doi:10.1074/jbc.M117.798850
50. Stinson BM, Moreno AT, Walter JC & Loparo JJ A Mechanism to Minimize Errors during Non-homologous End Joining. *Mol. Cell* 77, 1080–1091.e8 (2020). [PubMed: 31862156]
51. Kysela B et al. Ku stimulation of DNA ligase IV-dependent ligation requires inward movement along the DNA molecule. *J. Biol. Chem* 278, 22466–22474 (2003). [PubMed: 12682039]
52. Goodarzi AA & Lees-Miller SP Biochemical characterization of the ataxia-telangiectasia mutated (ATM) protein from human cells. *DNA Repair (Amst)* 3, 753–767 (2004). [PubMed: 15177184]
53. Yu Y et al. DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair (Amst)* 7, 1680–1692 (2008). [PubMed: 18644470]
54. Han Y, Reyes AA, Malik S & He Y Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature* 579, 452–455 (2020). [PubMed: 32188938]
55. Patel A, Toso D, Litvak A & Nogales E Efficient graphene oxide coating improves cryo-EM sample preparation and data collection from tilted grids. *BioRxiv* 1–19 (2021). doi:10.1101/2021.03.08.434344
56. Suloway C et al. Automated molecular microscopy: the new Legimon system. *J. Struct. Biol* 151, 41–60 (2005). [PubMed: 15890530]
57. Mastronarde DN Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol* 152, 36–51 (2005). [PubMed: 16182563]
58. Lander GC et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol* 166, 95–102 (2009). [PubMed: 19263523]

59. Voss NR, Yoshioka CK, Radermacher M, Potter CS & Carragher B DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol* 166, 205–213 (2009). [PubMed: 19374019]
60. Mindell JA & Grigorieff N Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol* 142, 334–347 (2003). [PubMed: 12781660]
61. van Heel M, Harauz G, Orlova EV, Schmidt R & Schatz M A new generation of the IMAGIC image processing system. *J. Struct. Biol* 116, 17–24 (1996). [PubMed: 8742718]
62. Tang G et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol* 157, 38–46 (2007). [PubMed: 16859925]
63. Punjani A, Rubinstein JL, Fleet DJ & Brubaker MA cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Meth* 14, 290–296 (2017).
64. Kimanius D, Forsberg BO, Scheres SH & Lindahl E Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* 5, 19 (2016).
65. Zhang K Gctf: Real-time CTF determination and correction. *J. Struct. Biol* 193, 1–12 (2016). [PubMed: 26592709]
66. Rohou A & Grigorieff N CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol* 192, 216–221 (2015). [PubMed: 26278980]
67. Scheres SHW & Chen S Prevention of overfitting in cryo-EM structure determination. *Nat Meth* 9, 853–854 (2012).
68. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7, 163 (2018).
69. Kucukelbir A, Sigworth FJ & Tagare HD Quantifying the local resolution of cryo-EM density maps. *Nat Meth* 11, 63–65 (2014).
70. Pettersen EF et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605–1612 (2004). [PubMed: 15264254]
71. Sánchez-García R et al. DeepEMhacer: a deep learning solution for cryo-EM volume post-processing. *24*, 1181 (2020).
72. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr* 66, 486–501 (2010). [PubMed: 20383002]
73. Croll TI ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* 74, 519–530 (2018). [PubMed: 29872003]
74. Saltzberg DJ et al. SSEThread: Integrative threading of the DNA-PKcs sequence based on data from chemical cross-linking and hydrogen deuterium exchange. *Prog. Biophys. Mol. Biol* 147, 92–102 (2019). [PubMed: 31570166]
75. Hepburn M et al. The Active DNA-PK Holoenzyme Occupies a Tensed State in a Staggered Synaptic Complex. *SSRN Journal* (2020). doi:10.2139/ssrn.3707279
76. Yang J et al. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A* 117, 1496–1503 (2020). [PubMed: 31896580]
77. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* 74, 531–544 (2018). [PubMed: 29872004]
78. Chang HHY, Watanabe G & Lieber MR Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent artemis resection at blunt DNA ends. *J. Biol. Chem* 290, 24036–24050 (2015). [PubMed: 26276388]

Figure 1. Cryo-EM structure of the LR synaptic complex.

a, Front (left) and top (right) views of the cryo-EM composite map (see Methods) of the LR complex assembled in the presence of ADP. **b**, Corresponding views of the structural model of the LR complex. Subunits are colored here and in all subsequent figures, as in **a**. **c**, An overview of the two DSBs within the LR complex. DNA elements are shown as solid ribbons and others in transparent representation. The distance between the two DSB ends is highlighted. **d**, Close-up view showing the interface between the LigIV N-BRCT (ribbon) and Ku70/80 core (surface) domains. Regions of the N-BRCT domain involving in the interaction are depicted. **e**, Close-up view showing the interface between the two copies of DNA-PKcs.

Figure 2. Activation mechanism of DNA-PKcs in the LR synaptic complex.

a, Superimposition of the DNA-PK complex in the apo states (left, PDB: 6ZHA³⁰; right, PDB: 7K0Y³⁹) with active DNA-PK derived from the LR complex. Transitions from the stand-alone states (gray) to the activated state are shown as curved arrows, indicating conformational changes induced by its assembly into the LR complex. **b**, Stabilization of a DNA-end blocking (DEB) helix within DNA-PKcs. Superimpositions of loops mediating DNA-PKcs–DNA-PKcs interaction, as well as the DNA end, are highlighted. **c**, Comparison of the kinase active site between the LR complex (olive), apo DNA-PK (purple³⁰ and red³⁹), mTOR (cyan⁴⁰), and CDK2 (green⁴¹). A flexible PIKK regulatory domain (PRD) represents the hallmark of activated kinase, as shown in all but DNA-PK in the LR complex. **d**, Structural organization of DNA-PKcs kinase active sites relative to different phosphorylation sites at the ABCDE and PQR cluster, as well as near the DNA-PK–DNA-PK and DNA-PKcs–Ku80 C-term helix interface.

Figure 3. Cryo-EM structure of the SR synaptic complex.

a, Front (left) and top (right) views of the cryo-EM composite map (see Methods) of the SR complex. **b**, Corresponding views of the structural model of the SR complex. **c**, Close-up view showing the interface between the LigIV DBD (ribbons) and Ku70 vWA (surface) domains. Regions of the LigIV DBD domain involved in the interaction are depicted. **d**, Close-up view showing the interface between the XLF CC (ribbons) and the two Ku70 vWA (surface) domains. Spheres depict the locations of missense or truncation mutation residues from cancer patients² that occur at the interface, although D176X and D178X truncations are also expected to impact XLF-Ku80 interaction due to the lack of XLF C-terminal KBM^{27,28}.

Figure 4, Structural transition from the LR to the SR synaptic complex.

a, Superimposition of the LR and SR complexes shown in front (left) and top (right) views. XLF homodimer is used for aligning the two conformers. DNA-PKcs and LigIV catalytic domains are hidden for clarity. The transition from the LR to the SR synaptic state indicates potential conformational changes induced by dissociation of DNA-PKcs and association of the LigIV catalytic domain. **b**, Model of structural transitions during NHEJ. After DSB detection by Ku70/80, DNA-PKcs is recruited to form a DNA-PK complex by inward translocation on DNA (1). A LigIV-XRCC4-XLF-XRCC4-LigIV scaffold assembles the two DNA-PK complexes into the LR complex, positioning the major auto-phosphorylation clusters near the kinase active centers in a *trans* manner (2). Upon auto-phosphorylation and dissociation of DNA-PKcs, the complex transitions into the SR complex, allowing the recognition and sealing of one nick by one LigIV catalytic domain (3). The complex next undergoes a conformational change and possibly DNA translocation to allow the recognition and sealing of the other nick by the second LigIV catalytic domain from the opposite side (4). The ligase factors then dissociate, and the DSB is repaired (5).